

ANALISIS *CLUSTER* DENGAN MENGGUNAKAN METODE *K-MEANS* UNTUK PENGELOMPOKKAN KABUPATEN/KOTA DI PROVINSI MALUKU BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN MANUSIA TAHUN 2014

M. W. Talakua¹, Z. A. Leleury², A. W. Talluta³

^{1,2,3}Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Pattimura
Jalan Ir. M. Putuhena, Kampus Unpatti, Poka, Ambon, Indonesia
e-mail: ¹ocat_14@yahoo.com; ³wandatalluta@yahoo.co.id

Abstrak

Pembangunan manusia merupakan suatu tujuan utama untuk mengukur keberhasilan sebuah negara. Salah satu aspek penting untuk mengukur tingkat pembangunan manusia adalah masyarakat yang unggul dari segi kuantitas dan kualitas, maka dilihat dari tiga dimensi kehidupan yaitu peluang hidup, pengetahuan, dan kehidupan yang layak. Penelitian ini membahas tentang pemanfaatan algoritma *k-means* untuk mengelompokkan kabupaten/kota di Provinsi Maluku berdasarkan kemiripan karakteristik daerah yang ditinjau dari lima ukuran Indeks Pembangunan Manusia (IPM). Lima ukuran tersebut adalah Angka Harapan Hidup (AHH), Angka Melek Huruf (AMH), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Per Kapita (PPK). Terdapat tiga *cluster* berdasarkan IPM yaitu: *cluster* 1: Kota Ambon, yang mempunyai angka IPM sangat maksimal. *Cluster* 2: MTB, Kepulauan Aru, SBB, SBT, MBD, dan Bursel, yang mempunyai angka IPM, AHH, AMH, RLS, dan PPK. *Cluster* 3: Malra, Malteng, Buru, Tual mempunyai angka IPM, AHH, AMH, RLS dan angka PPK. Berdasarkan angka Indeks Pembangunan Manusia, Angka Harapan Hidup, Angka Melek Huruf, Angka Rata-rata Lama Sekolah dan angka Pengeluaran Per Kapita, disimpulkan bahwa terdapat perbedaan yang signifikan pada tahun 2014.

Kata Kunci: analisis klaster, indeks pembangunan manusia, *k-means cluster*.

CLUSTER ANALYSIS BY USING K-MEANS METHOD FOR GROUPING OF DISTRICT/CITY IN MALUKU PROVINCE INDUSTRIAL BASED ON INDICATORS OF MALUKU DEVELOPMENT INDEX IN 2014

Abstract

Human development is a major goal to measure the success of a country. One important aspect to measure the level of human development is a society that is superior in terms of quantity and quality, it is seen from three dimension life that is the opportunity of life, knowledge, and a decent life. In this study discusses the utilization of *k-means* algorithm to cluster or grouping district or city in Maluku province. Based on the similarity of the regional characteristics in the review of the limits of the indicator size of IPM mortality of the area, namely Life Expectancy, Literacy Rate, Average School Attendance, and Income Per Capita. The purpose of this study is how to categorize and also to know whether there is a real difference between that groups. The data obtained from the central statistical agency then the results obtained from the five variables in telita there are three groups and also there are some differences found according to the characteristic of each variable. From this research, it becomes an input for the Maluku provincial government to be able to pay attention to all kinds of problems about IPM in this case the mortality of AHH, AMH, RLS and PPP. Based on human development index number, live expectancy, literacy rate, mean sof school duration and expendituri rate per capita, it can be conluded that there are significant differences in 2014

Keywords: Human Development Index, Cluster Analysis, K-Means Cluster

1. Pendahuluan

Pembangunan nasional Indonesia menempatkan rakyat sebagai titik sentral pembangunan. Untuk dapat ikut berpartisipasi dalam proses pembangunan, tentunya dibutuhkan masyarakat yang unggul dari segi kuantitas dan kualitas. Oleh karena itu, dirumuskan suatu konsep baru dalam mengukur pembangunan yang berorientasi pada manusia [1]. Indeks Pembangunan Manusia (IPM) merupakan pengukur perbandingan keberhasilan pembangunan manusia yang mencakup tiga dimensi kehidupan manusia. Diantaranya peluang hidup (diukur dengan indikator hidup yang sehat dan panjang umur yang diukur dengan harapan hidup saat kelahiran dalam tahun), pengetahuan (diukur dengan indikator angka tingkat baca tulis dan rata-rata lama sekolah dalam tahun), dan hidup layak (diukur dengan indikator logaritma natural dari produk domestik bruto per kapita dalam partisipasi daya beli per bulan dalam ribuan rupiah). Setiap tahun daftar negara menurut IPM diumumkan berdasarkan penilaian tersebut. Pengukuran alternatif lain adalah Indeks Kemiskinan Manusia yang lebih berfokus pada kemiskinan.

Pengelompokan wilayah kabupaten/kota di Maluku perlu dilakukan sebagai bahan perencanaan dan evaluasi sasaran program pemerintah. Tujuannya untuk meningkatkan angka pembangunan manusia berdasarkan indikator pembentuk IPM. Oleh karena itu, digunakan salah satu analisis statistik yaitu analisis *cluster* dengan metode *C-Means*. Analisa dilakukan dengan mengelompokkan kabupaten/kota di Provinsi Maluku berdasarkan indikator IPM. Indikator-indikator yang ada dinyatakan sebagai variabel dan IPM kabupaten/kota di Maluku dinyatakan sebagai objek yang dikelompokkan berdasarkan kemiripan karakteristiknya. Setiap objek lain berada dalam *cluster* yang sama.

Selain itu, salah satu algoritma data minim yang cukup populer digunakan baik dalam dunia bisnis, akademik, ataupun industri adalah Alogaritma *K-Means*. Alogaritma ini bekerja dengan cara membagi data dalam sejumlah *cluster* untuk dianalisis faktor kesamaan (*similarity*) maupun ketidaksamaan (*dissimilarity*) yang melekat pada kumpulan data tersebut. Kemudian dianalisis pola keterhubungan antar data.

Pada penelitian ini dibahas tentang pemanfaatan alogaritma *K-Means* untuk mengelompokkan kabupaten/kota di Provinsi Maluku berdasarkan kemiripan karakteristik daerah yang ditinjau dari empat ukuran indikator mortalitas IPM di daerah. Ukuran indikator tersebut adalah Angka Harapan Hidup (AHH), Angka Melek Huruf (AMH), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Per Kapita (PPK). Secara umum, *output* penelitian berupa hasil klasterisasi dapat digunakan sebagai sebuah acuan untuk menggambarkan distribusi pengelompokan berdasarkan kondisi profil IPM di Maluku.

2. Metodologi Penelitian

2.1 Analisis Multivariat

Analisis multivariat merupakan salah satu jenis analisis statistik yang digunakan untuk menganalisa data dengan banyak peubah bebas dan peubah terikat. Secara umum analisis multivariat di bagi menjadi dua metode, yaitu: metode dependen (*dependent method*) dan metode independen (*independent method*). Metode dependen digunakan untuk menganalisis ketergantungan di gunakan untuk menjelaskan atau memprediksi variabel terikat berdasarkan dua atau lebih variabel bebas. Sedangkan metode independen ini digunakan untuk menjelaskan seperangkat variabel untuk pengelompokan berdasarkan variabel-variabel tertentu. Metode ini dikelompokkan menjadi tiga, yaitu analisis faktor, analisis *cluster*, dan skala multidimensional.

Menurut Fadhli (2011), data dalam analisis multivariat dapat dinyatakan dalam bentuk matriks dengan n baris (objek) dan p kolom (variabel). Misalnya dalam suatu pengamatan terdapat p variabel, dari banyaknya n objek, maka data tersebut dapat di sajikan dalam bentuk matriks seperti berikut.

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1i} \\ X_{21} & X_{22} & \cdots & X_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{ni} \end{bmatrix}$$

2.2 Analisis Cluster

Clustering adalah proses membuat pengelompokan sehingga semua anggota dari setiap partisi mempunyai persamaan berdasarkan matriks tertentu. Analisis *cluster* atau analisis kelompok merupakan teknik analisa data yang bertujuan untuk mengelompokkan individu atau objek ke dalam beberapa kelompok yang memiliki sifat berbeda antar kelompok, sehingga individu atau objek yang terletak di dalam satu kelompok akan mempunyai sifat relatif homogen. Tujuan analisis *cluster* adalah mengelompokkan objek-objek tersebut.

Analisis *Cluster* memiliki beberapa kelebihan dan juga kekurangan sebagai berikut:

- 1) Kelebihan
 - a. Dapat mengelompokkan data observasi dalam jumlah besar dan variabel yang relatif banyak. Data yang di reduksi dengan kelompok akan mudah dianalisis.
 - b. Dapat dipakai dalam skala data ordinal, interval dan rasio.
- 2) Kelemahan
 - a. Pengelompokan bersifat subjektifitas peneliti karena hanya melihat dari gambar dendogram.
 - b. Untuk data heterogen antara objek penelitian yang satu dengan yang lain akan sulit bagi peneliti untuk menentukan jumlah kelompok yang di bentuk.
 - c. Metode-metode yang dipakai memberikan perbedaan yang signifikan, sehingga dalam perhitungan biasanya masing-masing metode dibandingkan.
 - d. Semakin besar observasi, biasanya tingkat kesalahan akan semakin besar.

2.3. Proses Analisis Cluster

- 1) Menetapkan ukuran jarak antar data.

Mengukur kesamaan antara objek. Sesuai prinsip dasar cluster yang mengeelompokkan objek yang mempunyai kemiripan, maka proses pertama adalah mengukur seberapa jauh ada kesamaan objek. Pengukuran jarak yang populer adalah metode *Euclidean distance*. Pada dasarnya, cara ini akan memasukan sebuah data ke dalam *cluster* tertentu dengan mengukur jarak data tersebut ke pusat *cluster*.

- 2) Melakukan proses standarisasi data jika diperlukan.
- 3) Melakukan pengklasteran.

Proses inti *clustering* adalah pengelompokan data, yang biasa dilakukan dengan dua metode yaitu:

- a) Metode Hierarki

Metode ini memulai pengelompokan dengan dua atau lebih objek yang mempunyai kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang mempunyai kedekatan kedua. Demikian seterusnya sehingga *cluster* akan membentuk semacam pohon dimana ada hierarki (tingkatan yang jelas) antara objek. Dendogram biasanya digunakan untuk membantu memperjelas proses hierarki tersebut.

- b) Metode Nonhierarki

Berbeda dengan metode hierarki, metode ini justru di mulai dengan menentukan terlebih dahulu jumlah *cluster* yang diinginkan. Setelah jumlah *cluster* diketahui, baru proses *cluster* dilakukan tanpa mengikuti proses hierarki. Metode ini juga disebut *K-Means cluster*.

- 4) Melakukan penanaman *cluster-cluster* yang terbentuk.
- 5) Melakukan validasi dan *profiling cluster*.

Adapun ciri-ciri *cluster* adalah homogenitas (kesamaan) yang tinggi antara anggota dalam satu *cluster* (*within cluster*) dan heterogenitas (perbedaan) yang tinggi antara *cluster* yang satu dengan *cluster* lain (*between-cluster*). Analisis cluster memiliki beberapa istilah penting, antara lain:

- 1) *Distances between cluster centers*, yaitu jarak yang menunjukkan bagaimana terpisahnya pasangan individu *cluster*.
- 2) Keanggotaan *cluster* (*cluster membership*), ialah keanggotaan yang menunjukkan *cluster* untuk setiap objek yang menjadi anggotanya
- 3) Pusat *cluster* (*cluster centers*), ialah titik awal di mulai pengelompokan dalam *cluster* nonhierarki.

- 4) Rata-rata lama *cluster* (*cluster centroid*), ialah nilai rata-rata variabel dari semua objek atau observasi dalam *cluster* tertentu.
- 5) Skedul aglomerasi (*agglomeration schedule*), ialah jadwal yang memberikan informasi tentang objek atau kasus yang dikelompokkan pada setiap tahap pada suatu proses analisis *cluster* yang hierarki.

2.4. Metode C-Means

C-Means merupakan salah satu metode data *clustering* nonhierarki yang berusaha mempartisi data yang ada dalam bentuk satu atau lebih *cluster*, sehingga data dengan karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama pula. Data dengan karakteristik yang berbeda dikelompokkan dalam kelompok yang lain. Metode yang termasuk dalam algoritma *clustering* berbasis jarak yang membagi data kedalam sejumlah *cluster* dan algoritma ini hanya bekerja pada atribut numerik.

C-Means merupakan salah satu algoritma dalam *data mining* yang biasa digunakan untuk melakukan *clustering* suatu data. Ada banyak pendekatan untuk membuat *cluster*, diantaranya adalah membuat aturan yang mendikte keanggotaan dalam kelompok yang sama berdasarkan tingkat persamaan di antara anggota-anggotanya. Pendekatan lainnya adalah dengan membuat sekumpulan fungsi yang mengukur beberapa properti dari pengelompokan tersebut sebagai fungsi dari beberapa parameter dari sebuah *clustering*. Mengelompokkan data dengan metode *C-Means* dilakukan dengan algoritma sebagai berikut.

- 1) Tentukan jumlah kelompok.
- 2) Alokasikan data ke dalam kelompok secara acak.
- 3) Hitung pusat kelompok dari data yang ada di masing-masing kelompok. Lokasi *centroid* setiap kelompok diambil dari rata-rata semua nilai data dari setiap fiturnya. Jika M menyatakan jumlah data dalam sebuah kelompok, i menyatakan fitur ke- i dalam sebuah kelompok, dan p menyatakan dimensi data, maka persamaan untuk menghitung *centroid* fitur ke- i digunakan persamaan (1).

$$C^1 = \frac{1}{M} \sum_{j=1}^M X_j \quad (1)$$

Persamaan (1) dilakukan sebanyak p dimensi dari $i = 1$ sampai dengan $i = p$.

- 4) Alokasikan masing-masing data ke *centroid*/rata-rata terdekat. Ada beberapa cara yang dapat dilakukan untuk mengukur jarak data ke pusat kelompok, diantaranya adalah Euclidean pengukuran jarak pada ruang jarak (*distance space*). Euclidean dapat dicari menggunakan persamaan 2.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2)$$

Pengalokasian kembali data ke dalam masing-masing kelompok dalam metode *K-Means* didasarkan pada perbandingan jarak antara data dengan *centroid* setiap kelompok yang ada. Data dialokasikan ulang secara tegas ke kelompok yang mempunyai *centroid* dengan jarak terdekat dari data tersebut. Pengalokasian data ini menurut MacQueen (1967) dapat ditentukan menggunakan persamaan (3).

$$a_{i1} = \begin{cases} 1, & d = \min\{D(X_i, C_l)\} \\ 0, & \text{lainnya} \end{cases} \quad (3)$$

a_{i1} adalah nilai keanggotaan titik x_i ke pusat kelompok C_l , d adalah jarak terpendek dari data x_i ke K kelompok setelah dibandingkan, dan C_l adalah *centroid* ke-1. Fungsi objektif yang digunakan untuk metode *K-Means* ditentukan berdasarkan jarak dan nilai keanggotaan data dalam kelompok. Fungsi objektif menurut MacQueen (1967) dapat ditentukan menggunakan persamaan (4).

$$J = \sum_{i=0}^n \sum_{l=1}^k a_{il} D(x_i, c_l)^2 \quad (4)$$

dengan n adalah jumlah data, k adalah jumlah kelompok, a_{i1} adalah nilai keanggotaan titik data x_i ke kelompok c_1 yang diikuti. a mempunyai nilai 0 atau 1. Apabila data merupakan anggota suatu kelompok, nilai $a_{i1} = 1$. Jika tidak, nilai $a_{i1} = 0$. Kembali ke langkah 3, apabila masih ada data yang berpindah kelompok atau apabila ada perubahan nilai *centroid* di atas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih di atas nilai ambang yang ditentukan.

2.5. Konsep Pembangunan Manusia

Manusia adalah kekayaan bangsa yang sesungguhnya. Tujuan utama dari pembangunan adalah menciptakan lingkungan memungkinkan bagi rakyatnya untuk menikmati umur panjang, sehat dan menjalankan kehidupan yang produktif. Hal ini tampaknya merupakan suatu kenyataan yang sederhana, akan tetapi sering terlupakan oleh berbagai kesibukan jangka pendek untuk mengumpulkan harta dan uang. Dalam [2], pembangunan manusia adalah suatu proses untuk memperbanyak pilihan-pilihan yang dimiliki manusia. Diantara banyak pilihan-pilihan tersebut, pilihan terpenting adalah untuk berumur panjang dan sehat, untuk berilmu pengetahuan, dan untuk mempunyai akses terhadap sumber daya yang dibutuhkan agar dapat hidup layak.

Pembangunan manusia adalah pembangunan yang berpusat pada manusia, yang menempatkan manusia sebagai tujuan akhir dan bukan hanya sebagai alat pembangunan. Konsep pembangunan manusia yang direkomendasikan oleh UNDP mencakup 4 indikator, yaitu kesejahteraan, produktifitas, pemberdayaan, dan berkelanjutan [1].

2.6. Indeks Pembangunan Manusia

Indeks Pembangunan Manusia (IPM) adalah pengukuran perbandingan dari harapan hidup, pendidikan dan standar hidup untuk semua negara diseluruh dunia. IPM digunakan untuk mengklasifikasikan apakah sebuah negara adalah negara maju, negara berkembang atau negara terbelakang, dan juga untuk mengukur pengaruh dari kebijakan ekonomi terhadap kualitas hidup. IPM mengukur pencapaian rata-rata sebuah negara dalam tiga dimensi dasar pembangunan manusia yaitu:

- 1) Hidup yang sehat dan panjang umur yang diukur dengan harapan hidup saat kelahiran.
- 2) Pengetahuan yang diukur dengan angka tingkat baca tulis dan rata-rata lama sekolah.
- 3) Standar kehidupan yang layak diukur dengan logaritma natural dari produk domestik bruto per kapita dalam paritas daya beli.

Terdapat empat indikator Indeks Pembangunan Manusia, yaitu (BPS 2012)

- 1) Angka Harapan Hidup
Angka Harapan Hidup (AHH) merupakan rata-rata perkiraan banyak tahun yang dapat ditempuh seseorang selama hidupnya. Perhitungan AHH melalui pendekatan tak langsung (*indirect estimation*). Jenis data yang digunakan adalah Anak Lahir hidup dan Anak Masih Hidup. Indeks harapan hidup dihitung dengan menghitung nilai maksimum dan nilai minimum harapan hidup sesuai standar UNDP, yaitu angka tertinggi sebagai batas atas untuk menghitung indeks dipakai 85 tahun dan terendah adalah 20 tahun.
- 2) Angka Melek Huruf
Angka Melek Huruf (AMH) adalah persentase penduduk usia 15 tahun ke atas yang bisa membaca dan menulis.
- 3) Rata-rata Lama Sekolah (RLS).
Rata-rata Lama Sekolah (RLS) adalah lama sekolah (tahun) penduduk usia 15 tahun ke atas.
- 4) Kemampuan daya beli penduduk yaitu kemampuan masyarakat secara ekonomi dalam memenuhi kebutuhan konsumsinya.

3. Hasil dan Pembahasan

3.1. Variabel Yang Digunakan

Data yang digunakan adalah data sekunder dari hasil pendataan Badan Pusat Statistik (BPS) Provinsi Maluku. Dalam penelitian ini variabel-variabel yang digunakan adalah indikator IPM sebagai berikut.

- X_1 : Persentase Angka IPM
- X_2 : Persentase AHH
- X_3 : Persentase AMH
- X_4 : Persentase Angka RLS
- X_5 : Persentase PPK.

3.2. Hasil Penelitian

3.2.1. Data indikator Indeks Pembangunan Manusia

Data yang digunakan dalam penelitian ini adalah data dari BPS tahun 2014, yakni IPM (X_1), AHH (X_2), AMH (X_3), RLS (X_4), dan PPK (X_5).

Tabel 1. Data Derajat Indikator IPM

No	Kabupaten/Kota	IPM	AHH	AMH	RLS	PPK
1	Maluku Tenggara Barat	59,81	62,50	11,81	8,88	5.600
2	Maluku Tenggara	62,74	63,98	12,06	8,71	6.804
3	Maluku Tengah	68,69	65,59	13,55	8,88	9.547
4	Buru	65,15	65,50	12,23	7,15	9.420
5	Kep. Aru	59,91	61,57	11,03	7,98	6.891
6	Seram Bagian Barat	62,39	59,90	12,70	8,42	7.779
7	Seram Bagian Timur	59,50	57,70	11,71	6,97	8.492
8	MBD	58,09	60,63	11,30	7,60	6.150
9	Buru selatan	60,74	65,36	11,69	6,62	6.915
10	Ambon	79,09	69,46	15,88	11,61	13.146
11	Tual	64,95	63,76	13,84	9,65	6.642

Data yang telah diperoleh dari BPS Provinsi Maluku, selanjutnya diproses menggunakan SPSS.16 untuk membentuk *cluster* berdasarkan indikator IPM di Provinsi Maluku dengan menggunakan metode nonhierarki.

3.2.2 Proses Pengelompokkan Dengan Metode Non Hierarki

Pada statistik deskriptif dibawah ini menunjukkan bahwa semua data lengkap yaitu angka IPM, AHH, AMH, angka RLS, dan angka PPK pada 11 kabupaten/kota di Provinsi Maluku.

Tabel 2. Statistik Deskriptif

	IPM	AHH	AMH	RLS	PPP
<i>N</i>	11	11	11	11	11
<i>Minimum</i>	58	57	11	6	5
<i>Maximum</i>	79	69	15	11	13
<i>Means</i>	63,18	62,64	12,00	7,73	7,44
<i>Std.Deviation</i>	6.080	3.355	1.265	1.421	2.255

3.2.3. Data Quick Cluster

Tabel 3. Hasil Output Data Quick Cluster SPSS.16

	<i>Initial Cluster Centers</i>		
	<i>Cluster</i>		
	1	2	3
IPM	2,60177	-0,68782	0,79249
AHH	1,89689	-1,68010	0,70456
AMH	2,31717	-0,79057	0,79057
RLS	2,30372	-1,21585	0,19198
PPP	2,46389	0,24631	0,68983

Output ini adalah tampilan pertama proses *clustering* data sebelum melakukan iterasi. Karena nanti akan dihasilkan proses *clustering* sesudah iterasi yang justru adalah hasil akhir *cluster*, maka *output* ini tidak akan dianalisis.

3.2.4. Proses Iterasi

Dari hasil *output* SPSS.16, dapat diketahui bahwa proses iterasi dilakukan sebanyak 3 kali. Proses ini dilakukan untuk mendapatkan *cluster* yang tepat dalam mengelompokkan 11 kabupaten/kota. Jarak minimum antar pusat *cluster* yang terjadi dari nilai minimum adalah 3,544.

Tabel 4. Final Cluster Centers

	<i>Final Cluster</i>		
	1	2	3
IPM	2,60177	-0,60558	0,25793
AHH	1,89689	-0,58713	0,40648
AMH	2,31717	-0,65881	0,39528
RLS	2,30372	-0,51194	0,19198
PPP	2,46389	-0,42702	0,02455

Berdasarkan *output* di Tabel 4, dapat diketahui bahwa data tersebut masih terkait dengan proses standarisasi yang mengacu pada *z-score* dengan ketentuan sebagai berikut.

- 1) Angka negatif (–) berarti data dibawa rata-rata total.
- 2) Angka positif (+) berarti diatas rata-rata total.

3.2.5. Tafsiran Setiap Cluster

Berdasarkan Tabel 3 dengan ketentuan yang telah dijelaskan, dapat didefinisikan sebagai berikut.

- Cluster 1:* Pada *cluster 1* diperoleh bahwa pada tahun 2014 angka RLS, angka PPK dan angka IPM dan AHH di atas rata-rata pada 11 kabupaten/kota di Provinsi Maluku. Angka RLS sangat minim. AMH dan angka PPK biasa dikatakan bahwa masih sedang diatas rata-rata. Sedangkan angka IPM dan AHH sangat banyak terjadi pada tahun 2014.
- Cluster 2:* Pada *cluster 2* diperoleh bahwa pada tahun 2014, angka RLS dan PPK di bawah rata-rata. AMH dapat dikatakan bahwa masih sedang dibandingkan dengan RLS dan PPK pada 11 kabupaten/kota di Provinsi Maluku. Dari ciri-ciri tersebut dapat dikatakan bahwa angka RLS dan angka PPK sangat rendah terjadi pada tahun 2014.
- Cluster 3:* Pada *cluster 3* diperoleh bahwa pada tahun 2014, AHH dan angka IPM di atas rata-rata sedangkan AMH masih sedang dibandingkan dengan angka RLS dan angka PPK di bawah rata-rata. Dari ciri tersebut, dapat dikatakan bahwa untuk angka PPK dan RLS sangat rendah terjadi pada tahun 2014.

Setelah membentuk 3 *cluster*, langkah selanjutnya adalah melihat apakah variabel-variabel yang telah terbentuk *cluster* mempunyai perbedaan pada setiap *cluster*. Dalam hal ini dapat dilihat dari *F* dan nilai probabilitas (*sig*) masing-masing variabel. Hal ini dilakukan dengan melihat *output* Anova berikut.

Tabel 5. Anova

ANOVA						
	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore(IPM)	4.618	2	.096	8	48.338	.000
Zscore(AHH)	3.164	2	.459	8	6.891	.018
Zscore(AMH)	4.427	2	.143	8	30.909	.000
Zscore(RLS)	3.514	2	.372	8	9.455	.008
Zscore(PPP)	3.584	2	.354	8	10.121	.006

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Pada Tabel 5 dapat dilihat bahwa kolom *cluster* menunjukkan besaran *between Cluster means*, sedangkan kolom *error* menunjukkan besaran *within Cluster means* sehingga nilai *F* diperoleh dari:

$$F = \frac{\text{between means}}{\text{within means}}$$

Sebagai contoh:

1. angka *F* pada *Z IPM* didapat dari:

$$F = \frac{\text{between means } Z \text{ IPM}}{\text{within means } Z \text{ IPM}}$$

$$F = \frac{4,618}{0,096} = 48,338$$

2. angka *F* = pada *Z AHH* didapat dari:

$$F = \frac{\text{between means } Z \text{ AHH}}{\text{within means } Z \text{ AHH}}$$

$$F = \frac{3,164}{0,459} = 6,891$$

3. angka F pada Z AMH didapat dari:

$$F = \frac{\text{between means Z AMH}}{\text{within means Z AMH}}$$

$$F = \frac{4,427}{0,143} = 30,909$$

4. angka F pada Z RLS didapat dari:

$$F = \frac{\text{between means Z RLS}}{\text{within means Z RLS}}$$

$$F = \frac{3,514}{0,372} = 9,455$$

5. angka F pada Z PPP didapat dari:

$$F = \frac{\text{between means Z PPP}}{\text{within means Z PPP}}$$

$$F = \frac{3,584}{0,354} = 10,121$$

Interprestasi angka F adalah semakin besar angka F suatu variabel dan angka signifikansinya adalah 0,05, maka semakin besar pula perbedaan variabel tersebut pada ketiga variabel. Sebagai contoh, angka F terbesar (48,338) ada pada Z IPM, dengan angka kolom SIG 0,000 yang berarti signifikansinya adalah nyata. Hal ini berarti faktor angka IPM sangat membedakan karakteristik ketiga *cluster* tersebut. Atau dapat dikatakan angka IPM pada ketiga *cluster* yang sangat berbeda antara *cluster* 1 dengan *cluster* lainnya. Perhatikan variabel Z AMH, yang mempunyai angka F sebesar 30,909 dan sig 0,00 yang berarti signifikansinya juga adalah nyata. Perhatikan pula perbedaannya dengan variabel Z AHH, yang mempunyai tingkat F sebesar 6,891 dan sig 0,18. Hal tersebut menyatakan bahwa signifikansinya di atas 0,05 (0,018 > 0,05). Maka variabel Z AHH pada *cluster* 1, *cluster* 2 dan *cluster* 3 tetap mempunyai perbedaan. Walaupun demikian, angka F variabel Z IPM lebih besar dari F variabel Z AMH. Dengan demikian, angka IPM lebih banyak dari angka melek huruf pada ketiga *cluster*.

3.2.6. Jumlah Anggota Setiap Cluster

Tabel 6. Number Of Cases In Each Cluster

Cluster	Jumlah Kab/Kota
1	1.000
2	6.000
3	4.000

Pada Tabel 6, dapat dilihat bahwa data terbanyak ada pada *cluster* 2, yaitu 6 kabupaten/kota. Data sedang ada pada *cluster* 3, yaitu 4 kabupaten/kota. Sedangkan data paling sedikit ada pada *cluster* 1 yaitu 1 kabupaten/kota. Karena tidak ada variabel yang hilang (*missing*), dengan demikian semua data sejumlah 11 kabupaten lengkap terdata pada ke 3 *cluster* dengan komposisi seperti di atas karena *cluster* 2 merupakan *cluster* terbesar.

3.2.7. Analisis Komposisi Cluster

Tabel 7. Komposisi Cluster

Kab/Kota	Cluster	Distance
MTB	2	1,053
MALUKU TENGGARA	3	0,945
MALUKU TENGAH	3	0,987
BURU	3	1,089
KEP. ARU	2	0,259
SERAM BAGIAN BARAT	2	1,182
SERAM BAGIAN TIMUR	2	1,472
MBD	2	0,404
BURU SELATAN	2	1,495
AMBON	1	0,000
TUAL	3	1,095

Contoh penafsiran:

- Cluster* 1: Kota Ambon mempunyai angka IPM dan AHH sangat maksimal (sangat banyak sedangkan jumlah AMH, angka RLS dan PPK berbanding balik dengan angka IPM yaitu terjadi pada tahun 2014.
- Cluster* 2: MTB, Kepulauan Aru, SBB, SBT, MBD, dan Buru Selatan mempunyai angka IPM, AHH, AMH, RLS, dan angka PPK sangat rendah terjadi pada tahun 2014.

- c) *Cluster 3*: Maluku Tenggara, Maluku Tengah, Buru, Tual mempunyai angka IPM, AHH, AMH, RLS, dan angka PPK sangat banyak terjadi pada tahun 2014.

4. Kesimpulan

Berdasarkan hasil dan pembahasan, maka dapat disimpulkan pengelompokan kabupaten/kota berdasarkan angka Indeks Pembangunan Manusia, Angka Harapan Hidup, Angka Melek Huruf, angka Rata-rata Lama Sekolah, dan angka Pengeluaran per Kapita kurang terjadi pada tahun 2014. Terdapat 3 *cluster* atau 3 kelompok yaitu:

- a) *Cluster 1* yakni Kota Ambon dengan angka IPM sangat maksimal. Sedangkan jumlah AHH, AMH, RLS, dan angka PPK sangat banyak terjadi pada tahun 2014.
- b) *Cluster 2* yakni MTB, Kepulauan Aru, SBB, SBT, MBD, dan Buru Selatan mempunyai angka IPM AHH, AMH, RLS, dan PPK kurang terjadi pada tahun 2014.
- c) *Cluster 3* : Malra, Malteng, Buru, Tual, mempunyai angka IPM, AHH, AMH, RLS dan angka PPK banyak terjadi pada tahun 2014.

Daftar Pustaka

- [1] Abdyo and H. Permadi, Metode Statistika Praktis, Malang: Universitas Negeri Malang, 1999.
- [2] B. P. S. P. Maluku, "Indikator Indeks Pembangunan Manusia 2010-2014," Badan Pusat Statistik Provinsi Maluku, Juli 2014. [Online]. Available: <https://maluku.bps.go.id/subject/26/indeks-pembangunan-manusia.html>. [Accessed Juli 2014].
- [3] B. P. S. P. Maluku, "Indikator Kesejahteraan Rakyat Provinsi Maluku 2014," Badan Pusat Statistik Provinsi Maluku, [Online]. Available: <https://maluku.bps.go.id/publication/2015/09/30/6cc18acb861407726a19addc/indikator-kesejahteraan-rakyat-provinsi-maluku-2014.html>.
- [4] Windha Mega P. D., "Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita," *Jurnal Informatika*, vol. 15, no. No.2, pp. 160-174, 2015.

